[Frontiers in Bioscience 8, d491-510, May 1, 2003]

VECTORS DERIVED FROM THE HUMAN IMMUNODEFICIENCY VIRUS, HIV-1

Edward Barker [1] and Vicente Planelles [2, 3]

[1] Department of Microbiology and Immunology, State University of New York, Upstate Medical University, 750 East Adams Street, Syracuse, NY, [2] Departments of Medicine and Microbiology and Immunology, University of Rochester Cancer Center, 601 Elmwood Avenue, Rochester, NY

FIGURES

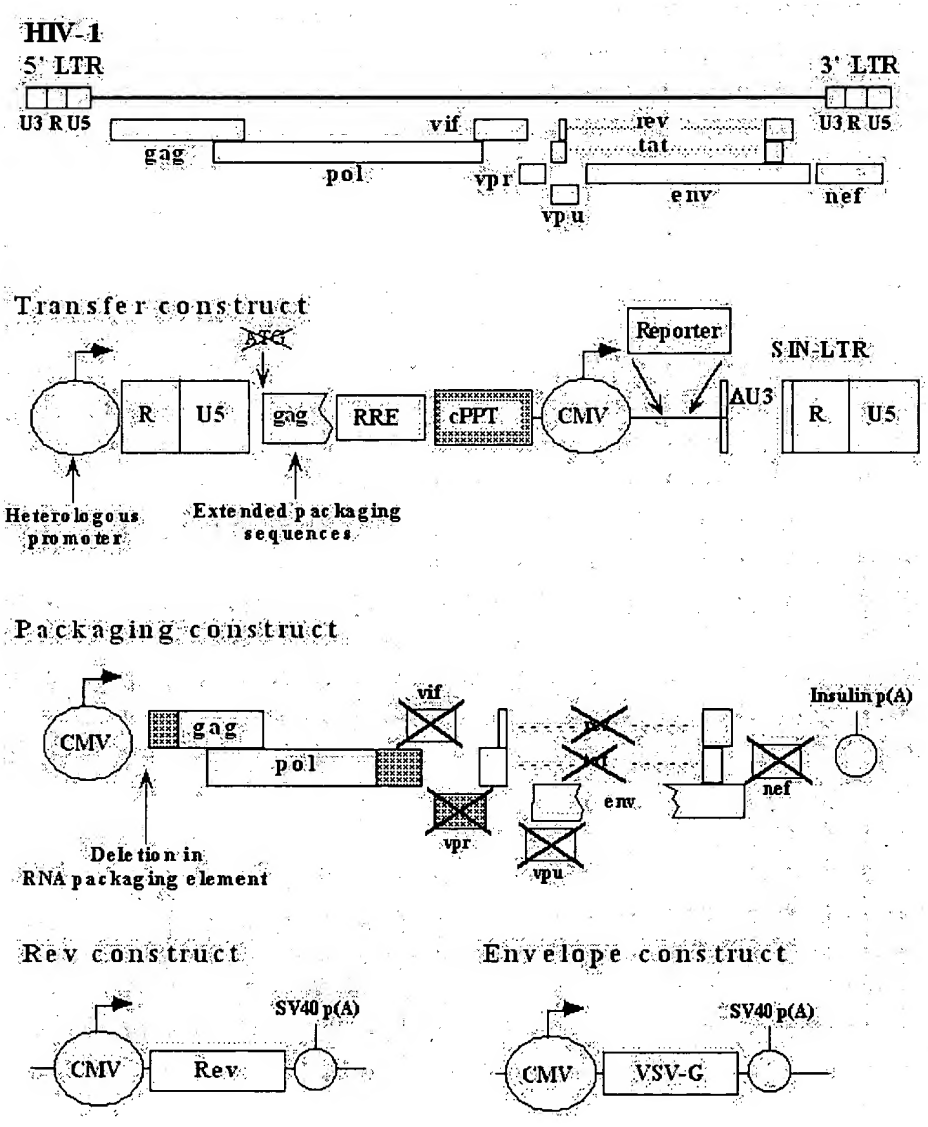

Figure 1. Schematic representation of the HIV-1 genome, and the genetic structure of a lentiviral vector.A latest generation lentiviral vector consists of four genetic elements, as shown. The transfer vector construct, is made up of SIN-LTRs in which the promoter/enhancer region from the 3' LTR (U3) have been deleted. The start site for *gag* is eliminated by frameshift mutation, and the coding sequence is truncated to only include the extended packaging sequences. The reporter gene is expressed from an internal promoter. The structural elements of the virion are provided as two sub-elements: a structural protein construct and a Rev construct. In the structural protein construct, RNA packaging sequence and the accessory genes (*vif, vpu, vpr* and *nef*) are deleted. The regulatory genes, *tat* and *rev* are also deleted. Shaded areas in the structural protein and transfer constructs denote determinants of infection of non-dividing cells. The envelope construct consists of a CMV driven expression cassette expressing VSV-G protein.

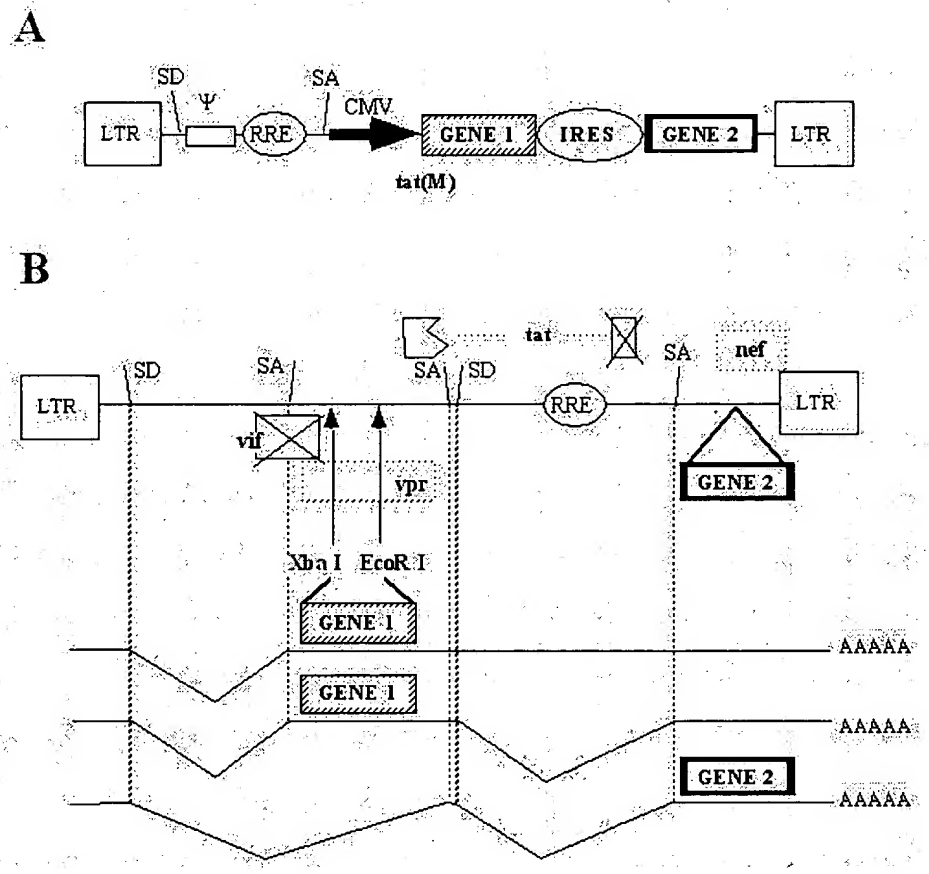

Figure 2. Design of multigene lentiviral vectors. A. Multigene vector based on the use of an internal ribosome entry site (IRES). B. Multigene vector based on natural HIV-1 splicing patterns. Discontinuous lines depict splicing events. SD, splice donor; SA, splice acceptor. Only splice donors and acceptors controlling expression of the reporter genes are shown. Mutations in the *tat* and *rev* open reading frames. The *rev* gene was inactivated by deletion of its second coding exon. The mutant, tat(M), was constructed by introducing a frameshift mutation after codon 24 of *tat*.

Interactions of INS (CRS) elements and the splicing machinery regulate the production of Rev-responsive mRNAs.

J Mol Biol. 1996 Mar 29;257(2):246-64. Unique Identifier : AIDSLINE MED/96180013

Mikaelian I; Krieg M; Gait MJ; Karn J; MRC Laboratory of Molecular Biology, Cambridge, UK.

Abstract: The human immunodeficiency virus type (HIV-1) Rev protein stimulates the export to the cytoplasm of unspliced HIV-1 mRNAs carrying the Rev response element (RRE). However, simple addition of the RRE to beta-globin pre-mRNA does not confer a Rev response on this heterologous transcript. In this paper, we demonstrate that a strong Rev response is conferred on beta-globin pre-mRNA when an inhibitory (INS) element is inserted into the gene together with the RRE. In the presence of INS element, Rev was able to stimulate the export to the cytoplasm of unspliced mRNA 10 to 15-fold. INS elements from the HIV-1 p17 gag and pol genes were equally active in complementing Rev-dependent nuclear export of unspliced mRNA. By contrast, mutated p17 gag INS element, known to be inactive in gag mRNA instability assays, was unable to complement the Rev/RRE system and stimulate nuclear export. Similarly, AUUUA-instability elements from the granulocyte-macrophage colony stimulating factor mRNA (GM-CSF) destabilised beta-globin mRNA but could not substitute for the HIV INS elements. Complementation between the Rev/RRE system and the INS elements was only observed when splicing was efficient. When splicing of the beta-globin gene receptor is impaired by mutations in the 5' splice donor, the 3' splice acceptor sequence, or the polypyrimidine tract, the majority of the unspliced mRNA is exported from the nucleus in the absence of Rev. In the presence of splice site mutations, Rev is able to act independently of a functional INS element and increase the export of unspliced mRNA three to fivefold. We propose that nuclear factor(s) binding to INS elements separate unspliced beta-globin pre-mRNA from the splicing apparatus. Pre-mRNA in this INS compartment remains accessible to Rev. Thus, there is a synergy between the INS elements and Rev which leads to enhanced nuclear export of unspliced mRNA.

Keywords: Animal Base Sequence Cell Nucleus/METABOLISM Cytoplasm/METABOLISM Gene Expression Gene Products, gag/GENETICS Gene Products, rev/*GENETICS Genes, gag/GENETICS Genes, pol/GENETICS Genes, Reporter/GENETICS Genetic Complementation Test Globin/GENETICS Granulocyte-Macrophage Colony-Stimulating Factor/GENETICS Hela Cells Human HIV Antigens/GENETICS HIV-1/*GENETICS/METABOLISM Models, Genetic Molecular Sequence Data Mutation Rabbits Regulatory Sequences, Nucleic Acid/*GENETICS RNA Precursors/*METABOLISM RNA Splicing/*GENETICS Support, Non-U.S. Gov't JOURNAL ARTICLE

960830

M9681175

MoBio Contents

RNA Splicing

Chapter 5 A|B|C|D|E

RNA splicing is a process that removes introns and joins exons in a primary transcript. An intron usually contains a clear signal for splicing (e.g., the beta globin gene). In some cases (e.g., the sex lethal gene of fruit fly), a splicing signal may be masked by a regulatory protein, resulting in **alternative splicing**. In rare cases (e.g., HIV genes), a pre-mRNA may contain several ambiguous splicing signals, resulting in a few alternatively spliced mRNAs.

Splicing signal

Most introns start from the sequence **GU** and end with the sequence **AG** (in the 5' to 3' direction). They are referred to as the **splice donor** and **splice acceptor** site, respectively. However, the sequences at the two sites are not sufficient to signal the presence of an intron. Another important sequence is called the **branch site** located 20 - 50 bases upstream of the acceptor site. The consensus sequence of the branch site is "CU(A/G)A(C/U)", where A is conserved in all genes.

In over 60% of cases, the exon sequence is (A/C)AG at the donor site, and G at the acceptor site.

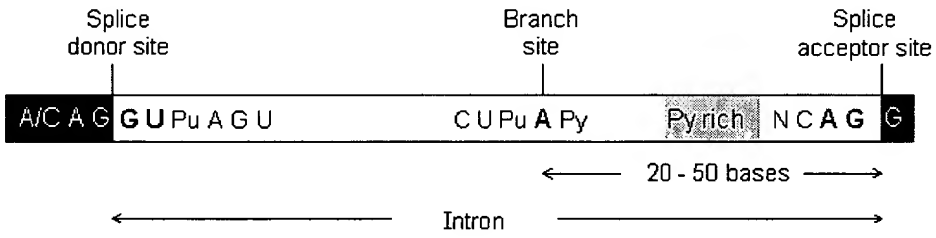

Figure 5-A-4. The consensus sequence for splicing. Pu = A or G; Py = C or U.

Splicing mechanism

The detailed splicing mechanism is quite complex. In short, it involves five snRNAs and their associated proteins. These ribonucleoproteins form a large (60S) complex, called **spliceosome**. Then, after a two-step enzymatic reaction, the intron is removed and two neighboring exons are joined together (see Alberts et al.). The branch point A residue plays a critical role in the enzymatic reaction.

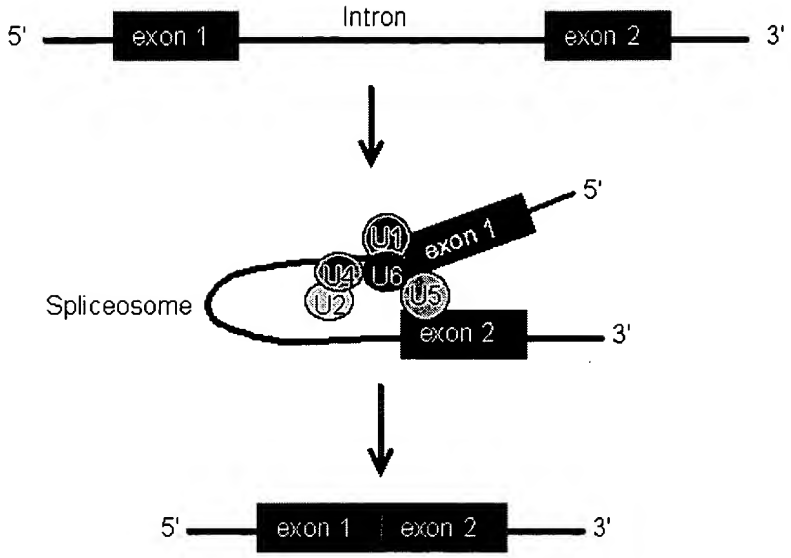

Figure 5-A-5. Schematic drawing for the formation of the spliceosome during RNA splicing. U1, U2, U4, U5 and U6 denote snRNAs and their associated proteins. The U3 snRNA is not involved in the RNA splicing, but is involved in the processing of pre-rRNA.

β-globin gene

Expression of the β-globin gene is a typical process. This gene contains two introns and three exons. Interestingly, the codon of the 30th amino acid, AGG, is separated by an intron. As a result, the first two nucleotides AG are in one exon and the third nucleotide G is in another exon.

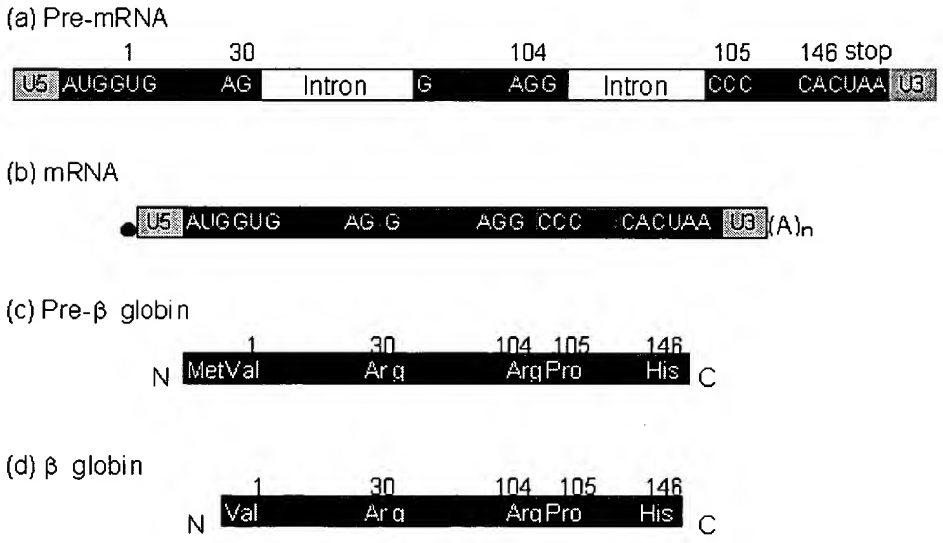

Figure 5-A-6. Expression of the human β-globin gene. U5 and U3 represent untranslated regions at the 5' and 3' end, respectively. Note that the mature β-globin protein does not contain the initiating methionine for protein synthesis.

Sex lethal gene

Sexual differentiation in *Drosophila* (fruit fly) is regulated by a protein called **sex-lethal (sxl)** protein. The female embryo expresses functional sxl proteins whereas the male embryo expresses non-functional sxl proteins. Their difference is a result of alternative splicing as shown in the following figure.

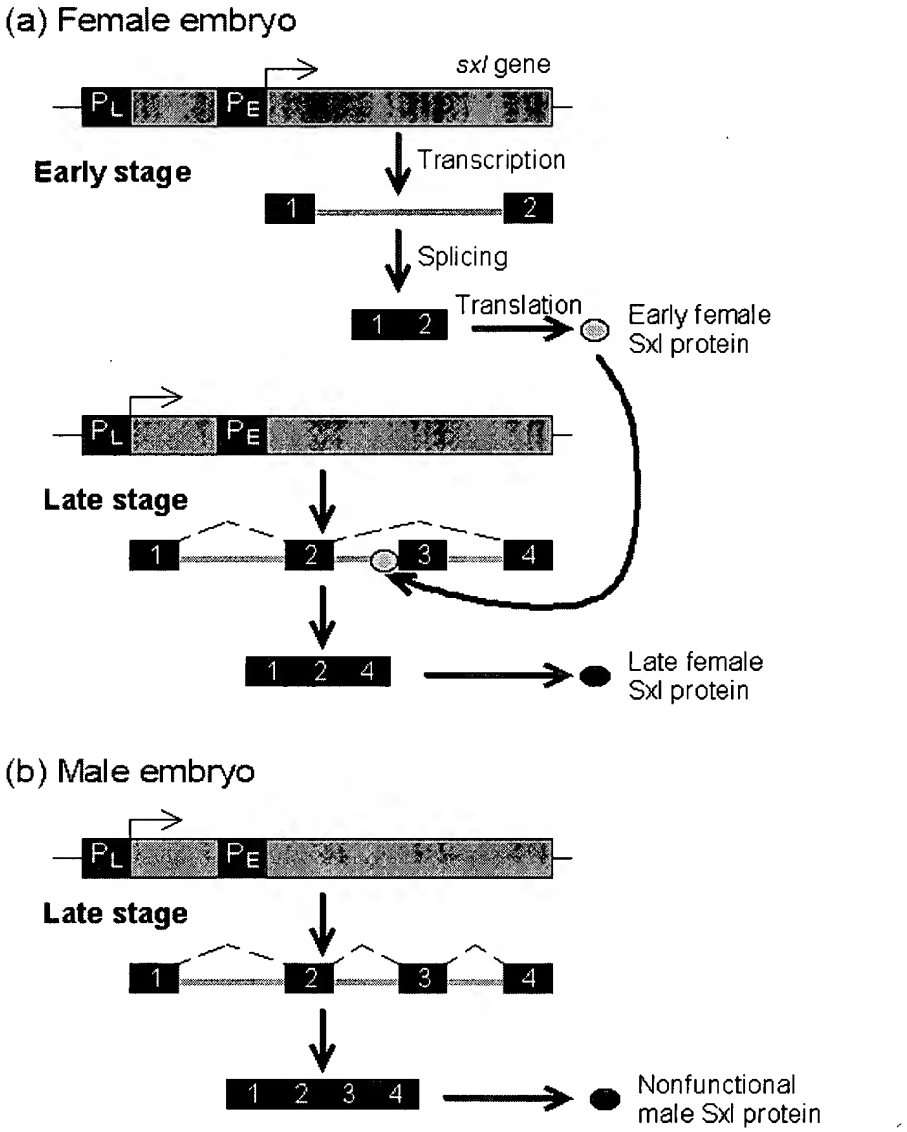

Figure 5-A-7. Expression of the *Drosophila* sex-lethal (sxl) protein.
(a) In the early stage of embryogenesis, the sxl protein is expressed in female embryo, but not in the male embryo.
(b) In the late female embryo, the sxl protein produced in the early stage may mask the splicing signal for the second intron, resulting in a different protein than in the male embryo.

The gene which encodes the sxl protein contains two promoters, denoted by P_L and P_E. P_L is active in the late development of both female and male embryos, but P_E is active only in the early stage of female embryogenesis. Therefore, in early embryogenesis, the sxl protein is expressed only in the female embryo.

The primary transcript generated by P_L consists of four exons separated by three introns. In the male embryo, the three introns are removed and all four exons are joined together. Its product is a non-functional sxl protein. In the female embryo, the sxl protein produced at the early stage may bind to the splice acceptor site of the second intron. As a result, the splicing machinery takes the next acceptor site for splicing. The third exon is then skipped, producing a functional sxl protein.

Exon skipping is also frequently observed when a critical residue in the splicing signal is mutated (example).

HIV-1 genome

The HIV-1 genome contains nine major genes: *gag, pol, vif, vpr, vpu, env, nef, rev* and *tat*. Their protein products are all derived from a single primary transcript. This is achieved by three mechanisms: (i) alternative splicing, (ii) leaky scanning of the initiation codon, and (iii) ribosomal frameshifting.

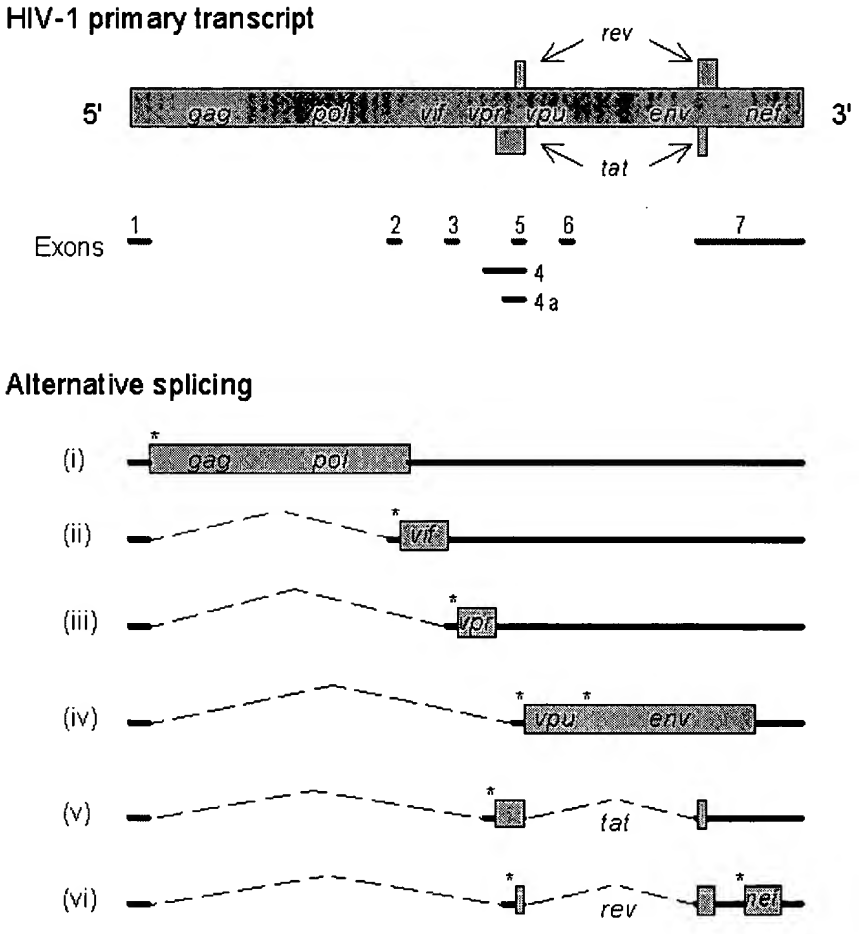

Figure 5-A-8. Alternative splicing of the HIV-1 primary transcript. (i) is unspliced, (ii) to (iv) are singly spliced, (v) and (vi) are doubly spliced. The resulting mRNA (i), (iv) and (vi) are bicistronic. The star "*" indicates the location of the initiation codon (AUG).

The HIV genome contains several ambiguous splicing signals, resulting in a few alternatively spliced mRNAs. They can be divided into three groups: (I) unspliced, (II) singly spliced, and (III) doubly spliced. As shown in the above figure, the resulting mRNA (i), (iv) and (vi) are **bicistronic** (each encoding two proteins). mRNA (i) encodes gag and pol proteins, mRNA (iv) encodes vpu and env, mRNA (vi) encodes rev and nef.

Protein synthesis starts from the initiation codon (AUG) and ends with one of three stop codons. In HIV, mRNA (iv) and (vi) have two initiation codons, but the first is sometimes skipped so that the second protein may be synthesized. mRNA (i) has only one initiation codon. Synthesis of the second protein (pol) is due to translational frameshifting (web link).